Biol Philos (2013) 28:577–592
DOI 10.1007/s10539-013-9374-2

Selection never dominates drift (nor vice versa)

Hayley Clatterbuck · Elliott Sober · Richard Lewontin

Received: 12 September 2012 / Accepted: 27 March 2013 / Published online: 3 May 2013

Abstract The probability that the fitter of two alleles will increase in frequency in a population goes up as the product of N (the effective population size) and s (the selection coefficient) increases. Discovering the distribution of values for this product across different alleles in different populations is a very important biological task. However, biologists often use the product Ns to define a different concept; they say that drift "dominates" selection or that drift is "stronger than" selection when Ns is much smaller than some threshold quantity (e.g., ½) and that the reverse is true when Ns is much larger than that threshold. We argue that the question of whether drift dominates selection for a single allele in a single population makes no sense. Selection and drift are causes of evolution, but there is no fact of the matter as to which cause is stronger in the evolution of any given allele.

Keywords Causal strength · Drift · Evolution · Neutrality · Selection

Pseudoproblems about causation

In that annus mirabilis 2007, a new diet plan was revealed to an avid public (Rolls 2007). It advises would-be weight losers to regulate what they eat by attending to "energy density." Big salads dressed with lemon juice have low energy density; candy bars have high. Energy density is measured by the ratio of calories to volume. A reporter at the *New York Times* offered the following summary: "you should eat things that have more weight than calories" (Schillinger 2007).

H. Clatterbuck (✉) · E. Sober
University of Wisconsin, Madison, WI, USA
e-mail: clatterbuck@wisc.edu

R. Lewontin
Harvard University, Cambridge, MA, USA

The previous paragraph mixes sense and nonsense. The ratio of calories to volume is less for the salad than it is for the candy bar. This comparison of the two foods makes sense. And let us grant, for the sake of argument, that a food's ratio of calories to volume affects its ability to promote weight loss. Volume matters indirectly; it promotes satiation and satiation promotes weight loss. Calories matter more directly. The lower the ratio of calories to volume is in the foods you eat, the higher the probability is that you'll lose weight—suppose that this is true. However, it makes no sense to ask, of a single food, whether it has more volume than calories. This is obvious. Volume and calories come in different units; they are incommensurable. But there is a second point that needs to be made: it makes no sense to ask, when you eat a diet of salads dressed with lemon juice for a week, whether your weight loss was caused more by the salad's volume than it was by its calorie count. This is a *pseudoproblem* (Carnap 1928).

This example would be a mere curiosity were it not for the fact that a similar issue arises in evolutionary biology. Drift and natural selection are causes of evolution.[1] If alleles differ in fitness (and so the selection coefficient s is nonzero) and the effective population size N is finite,[2] drift and selection jointly influence the evolutionary outcome. When a given allele increases in frequency, biologists often ask whether drift is "stronger than" (or "dominates") selection and answer this question by seeing whether the product Ns is much greater or much smaller than some threshold quantity (e.g., ½). We think, to the contrary, that the value of Ns does not provide a criterion for when drift is a stronger cause than selection. Further, we believe that serious problems arise when general methods for comparing causal strengths are applied to the case of drift and selection; this motivates our stronger conjecture that no measure of drift and selection can describe which cause is stronger. The question of whether drift dominates selection in the evolution of an allele is, we believe, a *pseudoproblem*.

Apportioning causal responsibility

Before we defend this thesis about selection and drift, some general remarks about causation are in order.

If A and B together cause E, how should one approach the question of which cause contributed more to that outcome? This question can be posed about event types or event tokens. In both cases, a natural approach, when the events in question are dichotomous, is to begin by considering the following counterfactuals:

[1] There now is disagreement among philosophers about this claim. Some hold that neither causes evolution (Walsh 2000; Walsh et al. 2002), others hold that only selection is a cause (Brandon 2008), while still others hold that both are causes (Sober 1984; Millstein 2002; Stephens 2004; Reisman and Forber 2005; Shapiro and Sober 2007, McShea and Brandon 2010). We count ourselves in this last category, though the main point of the present paper is not to criticize alternative positions.

[2] The effective population size (usually represented by the symbol "N_e") is the same as the census size when various assumptions hold (for example, random mating and equal numbers of males and females). The technical details of how the two quantities are related (see Charlesworth 2009 for details) won't matter in what follows.

- If A had occurred and B had not, what would the probability of E have been?
- If B had occurred and A had not, what would the probability of E have been?

If A and B are each necessary for the occurrence of E, there is a symmetry—the causal contributions are equal. If neither is necessary, their contributions may be unequal; if A makes a larger difference in E's probability than B does (when each cause is wiggled while the other is held fixed), then A makes the larger causal contribution.[3]

What if the causal variables are not dichotomous? For example, consider the fact that smoking cigarettes and inhaling asbestos both causally contribute to the occurrence of lung cancer. Increasing each while holding fixed the other raises the probability of the cancer. If someone smokes i cigarettes and inhales j milligrams of asbestos, and these together cause the person to get lung cancer, what would it mean for one of those causes to have made a larger contribution to that effect?

To consider what would happen if we varied each cause while holding fixed the other, we first need to make some choices. If the person had not smoked i cigarettes, how many would she have smoked? And if the person had not inhaled j milligrams of asbestos, how many milligrams would she have inhaled? Perhaps there is no such thing as *the* number of cigarettes the person in question would have smoked had she not smoked i (and ditto for asbestos dosage). Instead, there may be a frequency distribution of different amounts of cigarettes smoked and asbestos inhaled by individuals in the population, from which one can estimate the true probability distribution of smoking in the population; if so, the analysis might proceed by considering the expected number of cigarettes she would have smoked had she not smoked i (and similarly for the asbestos counterfactual). In some contexts, this procedure might be satisfactory, but in others it will not be. For example, there may be no probability distribution over these counterfactual dosages from which one can compute the expected value.[4] Whenever such troubles arise, the questions can be reformulated. We can make explicit what the counterfactual dosages are that we wish to consider:

- If the person had smoked h cigarettes rather than i (holding fixed the milligrams of asbestos she inhaled), what would the probability have been that she would get lung cancer?
- If the person had inhaled k milligrams of asbestos rather than j (holding fixed the number of cigarettes she smoked), what would the probability have been that she would get lung cancer?

These questions may have determinate answers, once values for h and k are chosen. However, the choice of values is often arbitrary. In such cases, it is arbitrary to say that smoking made a larger contribution than asbestos, or that the reverse was true, or that the causal contributions were equal.

[3] It is a further question whether "causal difference making" should be calibrated by arithmetic differences or by ratios; see Fitelson and Hitchcock (2011) for discussion.

[4] Or (as pointed out by an anonymous referee), the distribution might be bimodal where the expected value is in the valley between two peaks and the actual value is identical with the expected value.

There is a technique for comparing causal contributions that avoids having to arbitrarily choose values for h and k. Instead of asking whether a single person's cancer was due more to her smoking or to her asbestos exposure, we can consider all the individuals in a containing population. They differ in their exposures to cigarettes and to asbestos, and with enough individuals in the different "treatment cells," we can estimate the probability of lung cancer, conditional on different dosages, by examining how often individuals in those treatment cells get cancer and then using this frequency data to estimate the true probability of getting cancer for each dosage.[5] However, these conditional probabilities of getting cancer given different dosages of cigarettes smoked and asbestos inhaled do not by themselves provide us with a measure of causal strength.

We need to compare a "unit change" in asbestos with a "unit change" in cigarettes—which makes the larger difference in the probability of lung cancer? What would it mean to say that a unit change in cigarettes is "the same" as a unit change in asbestos? Why would a change of x cigarettes smoked be the correct unit to compare to a change of y milligrams of asbestos inhaled (rather than $2y$, say)?[6] An expedient solution that scientists sometimes deploy is to consider the de facto range of cigarette smoking in the population (which goes from 0 cigarettes in a lifetime to n) and the range of asbestos inhaling in that same population (which goes from 0 lifetime milligrams to m) and then to construct frequency distributions over this range (plotting the frequency of individuals in the population that smoke 0 cigarettes, 1 cigarette, and so on). Each of these frequency distributions has a standard deviation. The comparison of the two causes is then made by asking whether changing cigarette consumption by one standard deviation has more of an effect on the probability of lung cancer than changing asbestos consumption by one standard deviation.[7] The causal comparison can be made if we assume that a standard deviation in the one is "the same" as a standard deviation of change in the other.

There is room to wonder whether this technique for comparing the causal strengths of the two causes is adequate. It obviously is specific to the population in question; another population may show a different pair of standard deviations, so the apportioning of causal responsibility between smoking and asbestos may be

[5] Here, we are envisioning a maximum likelihood estimate of the true probability of cancer from frequencies in the population, though actual epidemiological studies may introduce assumptions based on theoretical considerations about the underlying probability distribution.

[6] It might be objected here that the problem can be solved as follows: if x cigarettes and y grams of asbestos both contain the same weight of carcinogenic materials, then we should compare how much each change affects the probability of lung cancer. The problem with this suggestion is that asbestos and cigarette smoke may contain different carcinogenic ingredients and these may differ in their potencies. An ounce of one poison need not have the same impact as an ounce of another. The fact that it is possible to measure two causes by using a "common currency" does not mean that it is appropriate to do so. For example, suppose that performance on a math test is influenced by a person's years of education and by the minutes spent studying for the test. While these causes are both measured in units of time, there is no reason to think that the strengths of these two causes should be compared by considering the effect of an additional 30 minutes of prior education with an additional 30 minutes of time spent studying for the test.

[7] There is a complication: the impact of a unit change in cigarettes may depend on what asbestos dosage is considered, and vice versa. If so, one can compute the average effect of each.

different in the two populations (Lewontin 1974; Sober 1988; Wright et al. 1992). But if this relativity to population is acknowledged, what is wrong with this way of handling nondichotomous causes? We will not pursue this question further, but will assume for the sake of argument that it makes sense. However, we think it is important to recognize that the causal question has changed. We began by asking whether a particular individual's cancer was due more to her smoking than to her asbestos exposure. This individual-level question leaves it open that individuals may differ—for two individuals who have the cancer after each smoked and inhaled asbestos (but had different dosages of each), smoking may be the dominant cause for one of them while asbestos may be the dominant cause for the other. Shifting to the population-level question means that this possible difference among individuals is no longer addressed; with the question shifted, there is a single fact of the matter as to whether a unit change in asbestos exposure affects the probability of lung cancer more than a unit change in smoking in the population. This population-level question is different from the individual-level question.

Selection and drift

Before addressing the question of whether it makes sense to say that drift is a stronger cause than selection in the evolution of a given allele, some history is in order. The neutral theory of evolution began by focusing on cases in which $s = 0$. This is the case of "strict neutrality"; if there is no variation in fitness, there is no natural selection at all. The question was how often genes evolve by this process of pure drift. But then Kimura (1968, 1983), Ohta (1992), and others saw the importance of considering the idea of "near-neutrality." Rather than asking how often alleles evolve purely by drift, the new question was to ask how often selection is weak enough to permit drift to play the dominant role in influencing outcomes. But how to make this liberalized conception concrete? This is where the product Ns was introduced. Kimura (1968) proposed that the criterion for nearly neutral evolution should be $Ns < ½$. Li (1978) argued that "a more reasonable definition" would be $Ns < 1$, although he notes that focusing on the dichotomy between $Ns < 1$ and $Ns > 1$ is "somewhat arbitrary" (374, 379). Roughgarden (1979, pp. 6–78) proposed $Ns \ll ½$ when there is selection against a dominant gene, or against a recessive, or when the heterozygote is intermediate; Roughgarden proposed the criterion of $Ns \ll ¼$ when there is heterozygote superiority. The transition from single "$<$" to double "$\ll$" is significant. Most biologists now recognize a gray zone. The usual practice is to say that selection "dominates" drift when Ns is much greater than some specified number and that drift "dominates" selection when Ns is much less than that number. In between, the two causes are said to be more coequal.

We think that describing the distribution of the quantity Ns across different alleles in different populations is a very important biological task. However, we see no justification for thinking that there is a uniquely correct criterion for "near neutrality." Still less do we think that the choice of this criterion permits one to say that selection is stronger than drift, or that the reverse is true, in the case of a single allele's evolution in a single population. What one can say about a population's

value for *Ns* is this: if the product were greater, the probability would be greater that the fitter of the two alleles increases in frequency, and if the product were smaller, that would raise the probability that the less fit allele increases in frequency.[8] Because the transitions are smooth and monotonic, there is no uniquely correct threshold that marks a boundary. But we want to defend a claim that goes beyond the line-drawing problem posed by the Sorites paradox: a given value for *Ns* (even an "extreme" one) doesn't tell you whether drift is stronger than selection. Our point is not epistemic. We are not saying that there is a fact of the matter about whether selection dominates drift and that scientists are cut off from knowing that fact. There is no such fact.[9]

How do the values of *N* and *s* for a given allele in a given population help predict how that allele will change in frequency? The two most prominent models of population dynamics under drift and selection, the Wright-Fisher model and its diffusion approximation, are probabilistic; they do not predict what must happen; rather, they describe different possible outcomes and assign a probability to each.[10] An illustrative example may be found in Fig. 1 (adapted from Roughgarden 1979, p. 78),[11] which depicts the case of balanced heterozygote superiority. Let the (normalized) fitness of the *Aa* genotype equal 1 and the fitnesses of *AA* and *aa* both have a value of 1-*s*. Suppose there is balanced mutational input; the probability of *A* mutating to *a* is *u* and that of *a* mutating to *A* is *v* (with both equal to 10^{-6} per generation).[12] If the process runs for a very long time, what are the probability densities of different allelic frequencies? Notice that all the curves are centered on $p = 0.5$; the variance around that mean value is smaller the larger *Ns* is.[13]

[8] This does not mean that if the product were smaller, the less fit allele would be more likely to evolve than the fitter allele.

[9] Matthen and Ariew (2002) make a similar claim: "Suppose that over a period of time a population stays exactly the same, or changes in some determinate way. The proposition that drift was involved to degree p in this history generally has no determinate truth value" (65). We agree with their conclusion but differ in our reasons. We take their argument to be that it does not make sense to talk about drift with respect to the births and deaths of token organisms or resulting changes to token populations; doing so would be a category mistake because drift only emerges when we consider types of populations. Our reasons for thinking that there is no fact of the matter regarding which cause was stronger persist even if we properly treat drift and selection as population-level causes.

[10] Most theoretical and philosophical discussions of drift assume the Wright-Fisher model. However, there are alternative models of population dynamics under drift and selection, and some of these models make different predictions about how drift and selection will interact (see Der et al. 2011 for a discussion). Our discussion in what follows will focus on the Wright-Fisher model.

[11] As an anonymous reviewer pointed out, this probability distribution is derived theoretically from a model. The probability distributions of lung cancer conditional on smoking and asbestos that we considered above were generated by inferring conditional probabilities from frequency data.

[12] We have not included mutation in our discussion until now. Doing so increases the realism of the example, but does not affect what we want to say about the comparison of selection and drift. Without mutation, the population frequencies of 0 and 1 in Fig. 1 would be absorbing states and there would be more weight at those points in the distribution.

[13] A similar pattern arises when we consider genetic set-ups different from heterozygote superiority. When there is selection for a dominant gene, or for a recessive gene, different values of *Ns* will correspond to different density curves. Curves with the same value of *s* but different values of *N* will have the same mean, but will differ in variance (Roughgarden 1979, pp. 77–78).

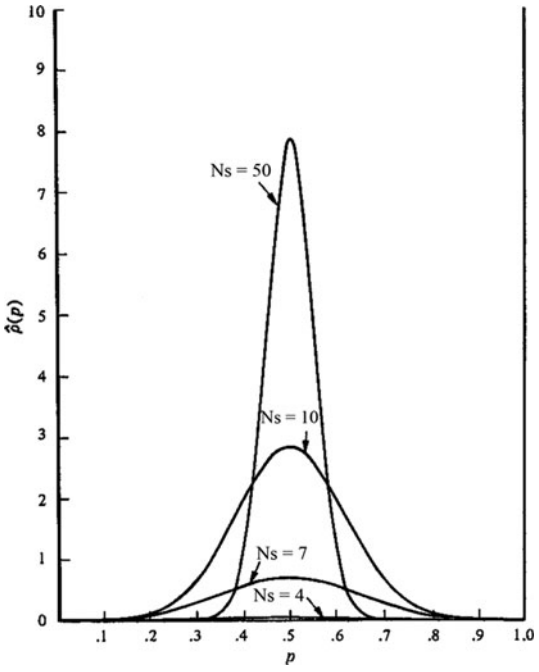

Fig. 1 The stationary ensemble distribution for different values of *Ns* with recurrent two-way mutation ($u = v = 10^{-6}$) and balanced heterozygote superiority ($w_{AA} = w_{aa} = 1 - s$ and $w_{Aa} = 1$). All curves are trimodal, with peaks at $p = 0$, $p = 0.5$ and $p = 1.0$; the first and last are too thin to be graphed (from Roughgarden 1979, p. 78)

How might this model be put to work in answering questions about causal contribution? Suppose that $N = \alpha$ and $s = \beta$ are the actual values of the two parameters and that the population has evolved for a very long time and now has a gene frequency of *f*. How probable would *f* have been if the values for *N* and *s* had been different? A shift from $N = \alpha$ to $N = \alpha^*$ will change the density curve (and so it will change the probability density of the observed value *f*). Ditto for a change from $s = \beta$ to $s = \beta^*$. The asterisked values can be chosen so that the first change makes more of a difference, or less, or the same, as you please. It is arbitrary to focus on comparisons that give drift the upper hand, or that do the same for selection. The fundamental point is that changing the value of the product *Ns* changes the shape of the density curve; that product can be manipulated by changing either or both of *N* and *s*.

What about the idea of looking at different alleles in different populations and seeing how *N* and *s* are each distributed? To determine whether selection or drift is a stronger cause in a population, perhaps one can examine the frequency distributions of the *N* and *s* values in populations "of the same kind" as the one we are considering. From this information, one can determine what a standard deviation of change in *N* and a standard deviation of change in *s* each is. Can one conclude from this that it makes sense to compare a unit change in one with a unit change in the other? Of course, there is the question of what the larger set of populations and

alleles is in which our example is located.[14] But more than that, it is important to see that this expedient cannot deliver the kind of judgment that those who talk about selection dominating drift (and vice versa) want to make. Biologists who talk of domination want it to make sense to say that drift dominates selection in this allele in this population while the reverse is true for the same allele in a different population or for a different allele in the same population. This heterogeneity among local judgments is not something that the technique we are now considering can deliver. It just hands you one big global thesis—that in the range of alleles and populations you are considering, a change of one standard deviation in selection coefficient makes more (or less) of a difference than a change of one standard deviation in effective population size.

The probability of fixation

In the curve we considered above, the product Ns determines the probability density distribution of allelic frequencies that holds after a great deal of time has passed. Therefore, for any manipulation of s, there is a corresponding change in N that will yield the same distribution. This symmetry in the contributions of selection and drift does not hold when we consider a special case of particular interest to biologists, the probability of fixation of a newly introduced allele.

In the Wright-Fisher model and its diffusion approximation from Kimura, the probability of fixation of an allele introduced at frequency $1/2\ N$ is approximately $2s$ (Wright 1931, 133; Kimura 1962, 715–716). Thus, it would appear that there is an evolutionary outcome for which changes in selection make more of a difference than changes in N (since N goes unmentioned in the $2s$ criterion), and therefore a case in which it makes sense to say that selection is a stronger determinant of that outcome. Appearances are misleading however, since an examination of the equations from which the $2s$ value is generated shows that s and N both play a role in determining the probability of fixation. From Wright (1931, 133),[15] the probability of fixation (π) of a beneficial allele introduced at frequency $1/2\ N$ is:

$$\pi = \frac{2s}{1 - e^{-4Ns}}$$

As the value of Ns increases, the denominator approaches 1, so $2s$ is a good approximation of the probability of fixation if $Ns > 1$. Once again, it is the product Ns that matters, so N and s make symmetrical contributions. It is true that manipulations of s will result in a greater change in the probability of fixation than manipulations of N as long as those manipulations keep Ns within the range of

[14] It is unclear which populations should be included in this data set. Should we focus on a specific allele and look at every population in which the same allele is present? Only those in which the allele has a non-zero selection coefficient? A positive coefficient? What about those populations from which the allele has been eliminated? Should we consider only conspecific populations or populations from a larger taxonomic unit?

[15] For a derivation and a more detailed discussion of the probability of fixation, see Kimura (1962, pp. 715–716).

values for which $2s$ is a good approximation. It might be tempting, then, to conclude that it takes a very large change in N to affect the probability of fixation and only a very small change in s, and that therefore selection is a stronger cause of fixation of a new mutation than drift is. However, this conclusion rests on the mistaken assumption that there is a non-arbitrary way of comparing unit changes in N and s.[16]

Zeroing-out

So far, we have contemplated counterfactuals that involve shifting from the actual values of N and s to two counterfactual values. A special case of this procedure involves "zeroing-out" each of the causes. We'll take zeroing-out selection to mean setting $s = 0$ and zeroing-out drift to mean setting $N = \infty$.[17] After exploring the consequences of this interpretation, we'll reflect on a complication.

Consider a gene that is selectively advantageous. Suppose the favored allele evolves to fixation in a finite population. Selection and drift are both present. Which is the stronger cause? What would happen if selection were occurring in a population in the limit as $N \to \infty$? In a haploid population in which there are two alleles A and B that have frequencies p and q and constant fitnesses $w(A)$ and $w(B)$, the change in allele frequency in the next generation is (Crow and Kimura 1970, p. 179):

$$\Delta p = \frac{spq}{\bar{w}}$$

The selection coefficient s represents the fitness difference $[w(A) - w(B)]$ and $\bar{w}$ is the average fitness of individuals in the parental generation. We are supposing that s is positive, so Δp is too. As p moves from 50 to 75 % and then on towards 100 %, the numerator shrinks and the denominator increases. In an infinite population, the favored allele never reaches 100 %; it steadily approaches that target as its rate of progress shrinks. Here evolution nods to Zeno. The fitter allele has no chance of reaching 100 % in an infinite population, though the probability of this outcome is positive if N is finite.[18]

[16] There are two additional features of this equation that are worth noting. Consider the following three possible solutions to the equation: (A) $Ns = 1$, $N = 1000$, $S = 0.001$, $\pi \approx 0.002$. (B) $Ns = 1$, $N = 100$, $S = 0.01$, $\pi \approx 0.02$, (C) $Ns = 0.1$, $N = 100$, $S = 0.001$, $\pi \approx 0.006$. Comparing A and B shows that the value of Ns does not fix the probability of fixation. Comparing A and C shows that reducing the value of Ns can increase the probability that the favorable allele will reach fixation. This last comparison may seem to contradict our earlier statement that increasing Ns increases the probability that the favored allele will increase in frequency; in fact, it does not. Increase in frequency is one thing; reaching fixation is another.

[17] Here and elsewhere, our discussion of what happens in "infinite" populations should be understood as shorthand for what happens as N approaches infinity. Discussing the frequency of a gene as $N \to \infty$ makes sense even if it makes no sense to ask what the frequency is of a gene in a population that is literally infinite.

[18] We have assumed that drift plays a causal role in a population if and only if the population is finite. Here we are using "drift" to name a causal property of a process, not to indicate a possible outcome of a process (e.g., deviation from an expected value). For helpful discussion of the distinction, see Stephens (2004).

Table 1 The probability of a gene's going to fixation, conditional on three possible conditions. Here $p_2 > p_1 > 0$ (Abrams 2007)

		Population size	
		Finite	Infinite
Selection coefficient	Positive	p_2	0
	Zero	p_1	

An anonymous referee pointed out that the fact that an event has a probability of zero does not entail that the occurrence of the event is impossible. We agree. Notice that our discussion of Table 1 involves just a comparison of probabilities

These facts are summarized in Table 1. If an advantageous gene goes to fixation in a finite population, the probability of that event is p_2. Had there been no selection (but the population remained the same size), the probability of fixation would have had the lesser (but still positive) value p_1. On the other hand, if there had been selection for the allele, but drift is zeroed-out by setting $N = \infty$, the probability of fixation is zero. So, when selection leads to fixation in a finite population, there being selection (as opposed to strict neutrality) makes less of a difference for the probability of fixation than there being finite population size (as opposed to infinite). In this case, the strengths of drift and selection are commensurable and the method zeroing-out entails that it is drift that always makes the bigger difference, since $(p_2 - 0) > (p_2 - p_1)$. It is curious that this result depends on nothing empirical. It has never been offered as a defense of the neutral theory of evolution; as noted earlier, that theory has focused on the empirical question of how values of *Ns* are distributed across alleles and populations.

A similarly odd result emerges when we consider outcomes other than fixation. Suppose we observe that the advantageous allele *A* is at 89 % in the population at time t_2 and ask whether drift or selection was the stronger cause of that outcome. To employ the zeroing-out strategy, we consider how probable that outcome would have been if there had been selection but no drift and how probable the outcome would have been had there been drift but no selection. These probabilities will be conditional on the state of the population at some earlier time t_1. Let's consider the probability that the population is at 89 % at t_2 given that it was at 63 % at time t_1.

The relevant considerations are depicted in Table 2. When there is selection but no drift, the starting frequency at t_1 says that there is one possible point frequency at t_2 that has a probability of one and that all the others have a probability of zero. So given that the population starts with a trait frequency of, say, 63 % at t_1, the probability is either one or zero that the trait frequency will be 89 % at t_2. When both selection and drift occur, all allelic states at t_2 are possible, so $p_4 > 0$. Now consider what the probability of the outcome would have been in a finite population with no selection. Again, from a starting state of 63 % *A*, all allelic states at t_2 have non-zero probabilities, so p_3 is positive, but its value will be smaller than p_4. Thus, except in the rare (=measure zero) case in which the observed state at t_2 is precisely the one that was determined to happen by the population's state at t_1, zeroing-out drift (moving from a finite to an infinite population) makes a bigger difference than

Table 2 The probability of a gene's being at 89 % at t_2, given that its frequency at t_1 was 63 %, in each of three possible conditions

		Population size	
		Finite	Infinite
Selection coefficient	Positive	p_4	0, 1
	Zero	p_3	

does zeroing-out selection, since $(p_4 - 0) > (p_4 - p_3)$. In the evolution of a favored allele, the zeroing-out strategy delivers the verdict that drift is stronger than selection in all but one possible case.[19]

We now consider the complication. In our running example of the person who smokes i cigarettes and inhales j grams of asbestos and then gets lung cancer, zeroing-out cigarettes means that the person smokes zero cigarettes. In that circumstance, it is obviously true that if she gets cancer, this cannot be due, even a little, to her smoking since she didn't smoke at all. But when you zero-out drift by setting $N = \infty$, it is still possible for the population to exhibit a trait frequency that differs from the one that is predicted by selection? What is true is that the population's probability of deviating from that predicted point frequency is zero. But events that have a probability of zero are not impossible. If an event that has a probability of zero does occur, what are we to say? Have we really zeroed-out drift? Perhaps not, since the population has just deviated from the point frequency predicted by selection. However, this interpretation means that it is impossible to zero-out drift. Drift (the cause) is calibrated by the value of N and the interpretation we are considering says that at every possible value, even when $N = \infty$, drift (the effect) can occur. Drift is now like a vampire that can't be killed even by driving a stake through its heart. We won't argue against this interpretation of zeroing-out, but we note that if zeroing-out drift really is impossible, then the method of zeroing-out cannot be used to ascertain whether selection is stronger than drift in a given case. On the other hand, if drift can be zeroed-out by setting $N = \infty$, then the method of zeroing-out entails that drift is almost always stronger than selection. This is a bizarre result in that it has nothing to do with the value of Ns and it is made doubly bizarre by the fact that it is obtainable a priori.[20]

A different approach

If Jack pushes a billiard ball east at the same time that Jill pushes it west, one can easily tell which push was stronger. Just see whether the ball moves east or west. This procedure does not require us to ponder counterfactuals: we don't need to ask

[19] A similar analysis can be carried out for the question of how selection and drift differentially affect the probability that a disadvantageous allele will increase in frequency.

[20] Okasha (2009, p. 28) endorses a zeroing-out strategy for determining the relative causal contributions of drift and selection. Such causal decomposition is possible, he argues, when we can identify the relevant causes and say, for each, what the effect would have been had the factor not been operative.

what Jack (Jill) would have done if he (she) had not pushed east (west). All we need to know is the empirical fact that Jack actually did one thing and Jill the other and there were no other causes at work that affected the trajectory. In this case, apportioning causal responsibility is easy.

This intuitively attractive approach can be used in evolutionary theory when drift is assumed to be zero. Consider, for example, the action of selection and mutation. Suppose selection favors allele *A* over allele *B* while the rate of mutation from *A* to *B* is greater than that from *B* to *A*. This last supposition means that there is mutation pressure that promotes the evolution of *B*. To tell whether selection for *A* is a stronger cause than the mutation pressure that promotes allele *B*, merely observe what happens in fact. If *A* increases in frequency, selection is stronger than mutation; if *A* declines, the reverse is true.

A similar analysis can be made when group selection and individual selection oppose each other, as they do when group selection favors altruism and individual selection favors selfishness. Group selection "pushes" altruism to increase in frequency; individual selection "pushes" selfishness to increase. If altruism increases in frequency in the metapopulation, group selection is the stronger cause; if altruism declines, then individual selection is stronger.

The previous analyses go smoothly under the assumption that the two causes together determine the magnitude and direction of the change that ensues. However, if population size is finite, we need to modify the protocol. Each cause is now conceived of in terms of the expected change in allele frequency that it entails; around each expected value there is a probability distribution. It is possible in this circumstance that an allele increases in frequency even though the cause that was pushing it to decline was stronger than the cause that was pushing it to increase. If the error distributions have small variance (because population size is large) and the allele frequency changes a lot, this possible mistake can be discounted.

When the evolution of an allele involves both selection and drift, can their joint impact on the observed change in allele frequency be disentangled by the strategy just described? If the gene in question is advantageous, we know that selection predicts that it should increase in frequency. But what does drift predict? It is not a vector pointing in the opposite direction. Rather, drift merely increases the variance of the probability distribution around the expected value that selection provides.[21] Observing an outcome that differs from the one that would occur if selection were the only cause doesn't mean that drift was the stronger cause. The magnitudes of selection and drift are given by the values of *N* and *s*. However, since the process thereby set in motion is stochastic, there are many possible outcomes. For fixed values of *N* and *s*, the fitter allele may increase in frequency, but it also is possible for it to decline. Suppose that we could make exact duplicates of a population, let

[21] This is not to say that drift processes never have a direction. For instance, we expect that a population with two alternate alleles, *A* and *a*, evolving under pure drift (with no mutational input) will eventually go to fixation for *A* or *a*. Which allele is expected to go to fixation is determined entirely by the starting frequencies of the two alleles. If the population starts at 80 % *A*, the probability of *A*'s going to fixation is 0.8, so we do expect that the population will move in that direction in the long term (Filler 2009). However, in a population evolving under both drift and selection, we still cannot say that if the population evolved in the direction predicted by drift alone that drift was therefore the stronger cause.

each duplicate "run", and observe how often the advantageous allele increases in frequency in this set of replicate populations. Let the populations run for as many generations as you please, and still, given enough trials, the fitter allele will increase in some trials and decrease in others.[22] However, by hypothesis, the causal strengths of drift and selection are identical in all these populations. Therefore, it would be a mistake to say that selection dominates drift when the favored allele increases in frequency and that the opposite is true when the fitter allele declines in frequency. This is not a mistake that is made by the conventional definition of domination (which views the value of *Ns* as settling the question of which cause is stronger). The conventional view holds that all the changes that ensue when *Ns* is greater than some threshold are to be described in one way while all the changes that ensue when *Ns* is less than the threshold should be described in another. The assumption is that the strengths of drift and selection should be understood in terms of features of the causal set-up, not in terms of which outcomes happen to ensue. We agree with this assumption, but think that the *Ns* criterion fails to deliver the goods.

Concluding comments

In denying that there are positive finite values of N and s that permit one to say that drift dominates selection in a population, or vice versa, we are not claiming that the values of N and s in real populations are uninteresting, or are irrelevant to predicting what will happen in those populations. Far from it. We have emphasized the biological importance of population size and strength of selection, since they jointly influence the probability of different allele frequencies obtaining.

We also are not denying that cases in which *Ns* is very small "resemble" cases in which s is zero, and that cases in which *Ns* is very large "resemble" cases in which N is infinite. But resemblance is always resemblance in some respect or other, and the fact that X resembles Y does not preclude X's also resembling Z, where Y and Z are markedly different. When *Ns* is very small, the density distribution for possible future allelic frequencies is close to the distribution entailed by the supposition that $s = 0$. For this reason, a simpler, drift-only, model may have a high degree of predictive accuracy even when the truth is that there is near-neutrality, not strict neutrality. But it is equally true that when *Ns* is small, the dynamics are even better approximated by N^*s^* where $s^* \gg s$ and $N^* \ll N$ (such that $N^*s^* = Ns$). With s^* large, selection is strong.

When *Ns* is very large, in what sense does this "resemble" cases in which N is infinite (and so there is no drift)? Well, the density distribution of possible future states is fairly well approximated by a simpler, selection-only model. But when N is infinite, selection cannot take advantageous traits to fixation, so in this respect the "resemblance" is weak. Furthermore, although a large *Ns* resembles (in some respects) a model in which N is infinite, it also resembles cases in which $N^* \ll N$ and

[22] A coin-flipping analogy is helpful here. If we flip a fair coin 5 times, we would not be too surprised to observe an outcome of 5 heads. If we flip a fair coin 1,000 times, it is much less probable that we will get 1,000 heads. However, it is still possible, and given enough coin-flipping experiments, the probability that we would eventually observe this outcome approaches unity.

$s^* \gg s$ (such that $N^*s^* = Ns$). So a large value for Ns resembles some cases in which population size is smaller, and so drift is strong.

For these reasons, the Ns criterion might serve as a useful rule of thumb for when a population can be accurately modeled by neutral, selection-only, or mixed models. We have no beef with this heuristic usage by biologists, but we do object to drawing causal conclusions from the resemblance of populations where Ns is very small to populations that contain no selection at all (or the resemblance of populations where Ns is very large to populations with no drift).

The relationship of selection to drift resembles the relationship of the probability a coin has of landing heads and the number of times the coin is tossed. If you toss a fair coin some number of times, the expected frequency of heads is 50 %. The fact that you tossed ten times rather than 1,000 affects the probability distribution you draw around that expected frequency. With ten tosses, there is a large probability that the coin's frequency of heads will be outside of the interval 40–60 %. Had you tossed 1,000 times, the chance of this deviation from the expected value would have been much smaller. Suppose you toss the fair coin ten times and the outcome is 40 % heads. This outcome was due to the fact that the coin was fair and to the fact that you tossed it ten times. Do not ask which cause was stronger. You can change the probability of that outcome by changing either the bias of the coin or the number of times it is tossed (Sober 1984, p. 117; Stephens 2004, p. 556).

In saying that drift and selection are both causes of evolution, we are not saying that they are "separate processes." We abstain from using this formulation because we don't wish to take a stand on how processes should be individuated. In the case of tossing a coin repeatedly, it is natural to say that there is just one process and that it has two characteristics—there is the bias of the coin and the number of times the coin is tossed. We are happy to use this format in describing concrete cases of evolution. All populations are finite, so every instance of allele evolution involves drift. When selection and drift both influence an allele's evolution, there is no need to locate each of them in its own autonomous "process."[23] It is natural to say that there is one process of evolution under way and that it has several facets. This all that is needed for drift and selection to both be causes of evolution.

Although drift and selection together cause alleles to evolve, it makes no sense to ask which cause was stronger in the evolution of a given allele. Or more precisely, this question makes no sense if the answer is required to satisfy the following constraints:

1. The answer for one allele in a given population may differ from the answer for the same allele in a different population, and also from a different allele in the same population.

[23] Beatty (1984), Hodge (1987), and Millstein (2002) embrace this "separate process view" when they define drift as an "indiscriminate sampling process" and selection as a "discriminate sampling process". This position entails that no single process can be both drift and selection, since a sampling process that is discriminating can't be indiscriminate, and vice versa. Sober and Shapiro (2007, p. 256) make the mistake of saying that selection and drift are distinct processes, a claim that was extraneous to their larger argument. For a defense of the view that drift and selection are different aspects of a single process, see Abrams (2007).

2. The answer does not depend on arbitrary choices concerning what counterfactual alternatives to the actual values of N and s one should consider.

In arguing for this point, we have taken "dominates" and "is stronger than" seriously. If one simply stipulates that "selection dominates drift" means that Ns is a lot bigger than some chosen quantity, there is no substantive question to discuss. But the relative strengths of two causes that jointly produce an effect should not be settled by adopting an arbitrary stipulation. Judged by that standard, selection and drift are both causes of an allele's evolution but there is no fact of the matter as to which was the stronger cause.[24]

Acknowledgments We are grateful to Martin Barrett, David Baum, Michael Goldsby, Daniel Hausman, Trevor Pearce, Reuben Stern, Elena Spitzer, Mike Steel, Naftali Weinberger, and to the anonymous referees of this journal for useful comments on an earlier draft.

References

Abrams M (2007) How do natural selection and random drift interact? Philos Sci 74:666–679
Beatty J (1984) Chance and natural selection. Philos Sci 51:183–211
Brandon R. (2008) Natural selection. In Zalta E (ed) Stanford Encycl Philos. http://plato.stanford.edu/entries/natural-selection/#NatSelDri. Accessed 15 Apr 2012
Carnap R (1928) Scheinprobleme in der philosophie. Weltkreis-Verlag, Berlin
Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Gen Rev 10:195–205
Crow J, Kimura M (1970) An introduction to population genetics theory. Burgess Publishing Co., Minneapolis
Der R, Epstein CL, Plotkin JB (2011) Generalized population models and the nature of genetic drift. Theor Popul Biol 80:80–99
Filler J (2009) Newtonian forces and evolutionary biology: a problem and solution for extending the force interpretation. Philos Sci 76:774–783
Fitelson B, Hitchcock C (2011) Probabilistic measures of causal strength. In: Illari P, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 600–627
Hodge J (1987) Natural selection as a causal, empirical, and probabilistic theory. In: Kruger L, Gigerenzer G, Morgan M (eds) The probabilistic revolution, vol 2., MIT PressCambridge, MA, pp 233–270
Kimura M (1962) On the probability of fixation of mutant genes in a population. Genet 47:713–719
Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet Res 11:247–269
Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, New York
Li WH (1978) Maintenance of genetic variability under the joint effect of mutation, selection, and random drift. Genet 90:349–382
Lewontin R (1974) The analysis of variance and the analysis of causes. Am J Hum Genet 25:400–441
Matthen M, Ariew A (2002) Two ways of thinking about fitness and natural selection. J Philos 99:55–83
McShea D, Brandon R (2010) Biology's first law. University of Chicago Press, Chicago
Millstein R (2002) Are random drift and natural selection conceptually distinct? Biol Philos 171:33–53
Ohta T (1992) The nearly neutral theory of molecular evolution. Ann Rev Ecol Syst 23:263–286

[24] Sober (2011) distinguishes three types of conventionalism about causes. Qualitative conventionalism about X and Y says that there is no fact of the matter as to whether X and Y both cause Z. Comparative conventionalism says that when X and Y both cause Z, there is no fact of the matter as to which is stronger. And quantitative conventionalism says that when X and Y both cause Z there is no fact of the matter as to how strong each cause is. Sober was discussing group and individual selection in that paper, and was describing alternatives to the three-fold realism of Okasha (2009), but the three-fold distinction and the associated triplet of realist positions pertain to the present topic of selection versus drift.

Okasha S (2009) Evolution and the levels of selection. Oxford University Press, New York
Rolls B (2007) The volumetrics eating plan. Harper, New York
Reisman K, Forber P (2005) Manipulation and the causes of evolution. Philos Sci 72:1115–1125
Roughgarden J (1979) Theory of population genetics and evolutionary ecology. Prentice-Hall, Upper Saddle River
Schillinger L (2007, July 22) Diets to, um, satisfy every appetite. New York Times, p ST9
Shapiro L, Sober E (2007) Epiphenomenalism—the do's and the don'ts. In: Wolters G, Machamer P (eds) Studies in causality: historical and contemporary. University of Pittsburgh Press, Pittsburgh, pp 235–264
Sober E (1984) The nature of selection. MIT Press, Cambridge
Sober E (1988) Apportioning casual responsibility. J Philos 85:303–318
Sober E (2011) Realism, conventionalism, and causal decomposition in units of selection. Philos Phenomenol Res 82:221–231
Stephens C (2004) Selection, drift, and the 'forces' of evolution. Philos Sci 71:550–570
Walsh D (2000) Chasing shadows—natural selection and adaptation. Stud Hist Philos Biol Biomed Sci 31:135–153
Walsh D, Lewens T, Ariew A (2002) The trials of life—natural selection and random drift. Philos Sci 69:452–473
Wright E, Levine A, Sober E (1992) Reconstructing Marxism—essays on explanation and the theory of history. Verso, New York
Wright S (1931) Evolution in Mendelian populations. Genet 16:97–159